CrimeBoard (Gathering, Scraping, Munging and Cleaning Data)

Class: DAMG6210

CRN: 16149

Team members:

Sayeed Ahmed - 002191535

Natarajan Lekshmi Narayana Pillai - 002766033

Chaman Betrabet – 002784662

Sunil Kumar Rudrakumar – 002764807

Files in ZIP

crimeboard_sample_data.zip – Sample data

CrimeBoard ER Diagram - Assignment 3.pdf – ER Diagram

EDA_CRIMEBOARD.ipynb – EDA

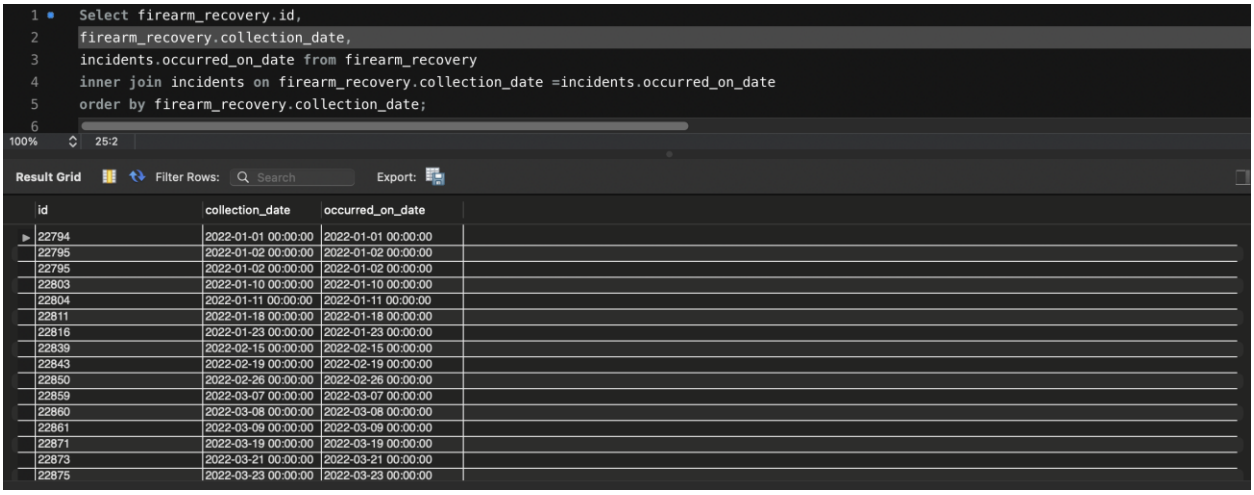
Crimeboard SQL Assignment 3.sql – SQL queries for Use cases

CrimeBoard - DataScraper for Assignment3.py – Python script for Gathering, Scraping, Munging and Cleaning Data

Use Cases

1. **Which firearm recovery incident date correlates with the incident report date?**

   Select firearm_recovery.id, firearm_recovery.collection_data, incidents.occured_on_date from firearm_recovery inner join incidents on firearm_recovery.collection_data =incidents.occured_on_date
   order by firearm_recovery.collection_data;



```
Select firearm_recovery.id,
firearm_recovery.collection_date,
incidents.occurred_on_date from firearm_recovery
inner join incidents on firearm_recovery.collection_date =incidents.occurred_on_date
order by firearm_recovery.collection_date;
```

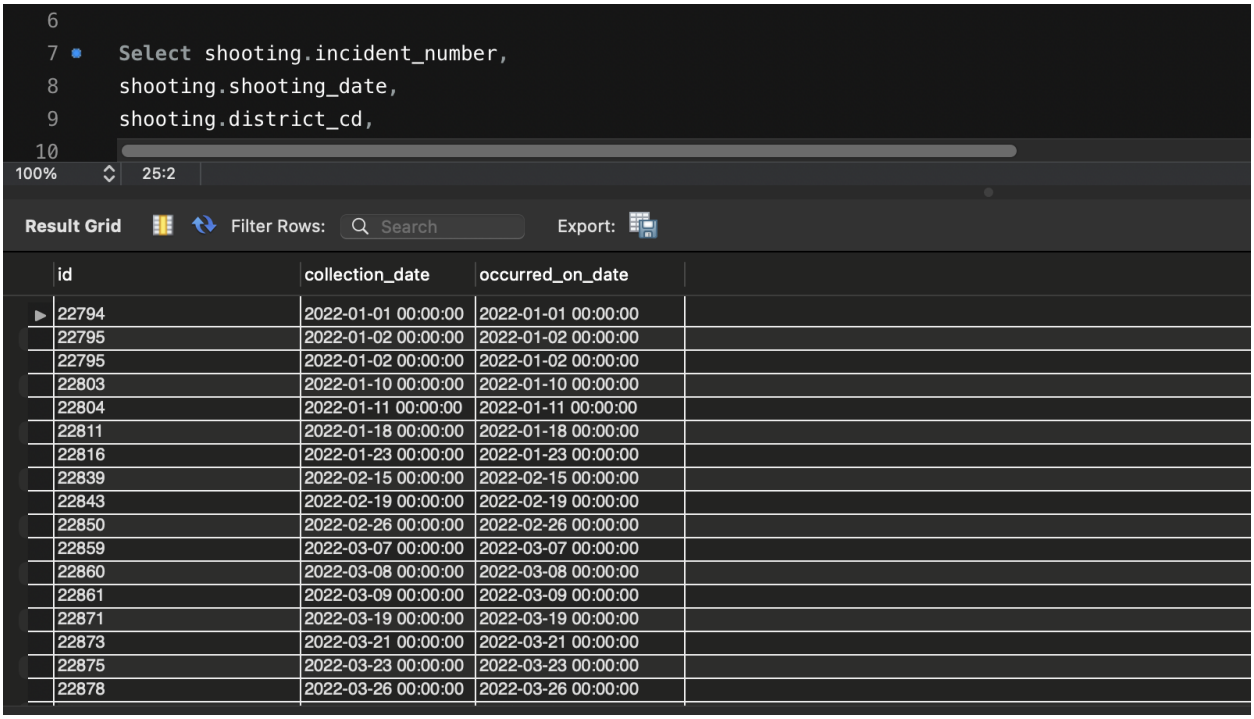
100% 25:2

Result Grid | Filter Rows: Search | Export:

id	collection_date	occurred_on_date
22794	2022-01-01 00:00:00	2022-01-01 00:00:00
22795	2022-01-02 00:00:00	2022-01-02 00:00:00
22795	2022-01-02 00:00:00	2022-01-02 00:00:00
22803	2022-01-10 00:00:00	2022-01-10 00:00:00
22804	2022-01-11 00:00:00	2022-01-11 00:00:00
22811	2022-01-18 00:00:00	2022-01-18 00:00:00
22816	2022-01-23 00:00:00	2022-01-23 00:00:00
22839	2022-02-15 00:00:00	2022-02-15 00:00:00
22843	2022-02-19 00:00:00	2022-02-19 00:00:00
22850	2022-02-26 00:00:00	2022-02-26 00:00:00
22859	2022-03-07 00:00:00	2022-03-07 00:00:00
22860	2022-03-08 00:00:00	2022-03-08 00:00:00
22861	2022-03-09 00:00:00	2022-03-09 00:00:00
22871	2022-03-19 00:00:00	2022-03-19 00:00:00
22873	2022-03-21 00:00:00	2022-03-21 00:00:00
22875	2022-03-23 00:00:00	2022-03-23 00:00:00

2. **Which shooting date correlates with the incident report date?**

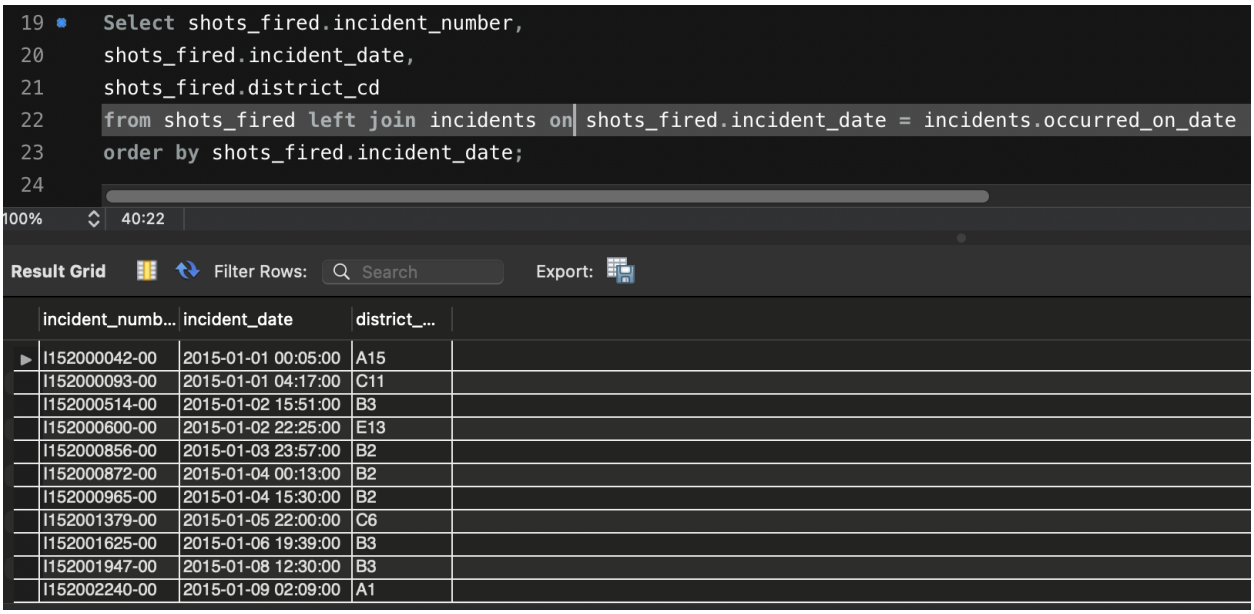
Select shooting.incident_number, shooting.shooting_date, shooting.district_cd, shooting.shooting_type, shooting.victim, shooting.victim_gender, shooting.victim_race, shooting.victim_ethnicity, shooting.multi_victim from shooting left join incidents on shooting.shooting_date = incidents.occurred_on_date
where incident.shooting = 1
order by shooting.shooting_date;



id	collection_date	occurred_on_date
22794	2022-01-01 00:00:00	2022-01-01 00:00:00
22795	2022-01-02 00:00:00	2022-01-02 00:00:00
22795	2022-01-02 00:00:00	2022-01-02 00:00:00
22803	2022-01-10 00:00:00	2022-01-10 00:00:00
22804	2022-01-11 00:00:00	2022-01-11 00:00:00
22811	2022-01-18 00:00:00	2022-01-18 00:00:00
22816	2022-01-23 00:00:00	2022-01-23 00:00:00
22839	2022-02-15 00:00:00	2022-02-15 00:00:00
22843	2022-02-19 00:00:00	2022-02-19 00:00:00
22850	2022-02-26 00:00:00	2022-02-26 00:00:00
22859	2022-03-07 00:00:00	2022-03-07 00:00:00
22860	2022-03-08 00:00:00	2022-03-08 00:00:00
22861	2022-03-09 00:00:00	2022-03-09 00:00:00
22871	2022-03-19 00:00:00	2022-03-19 00:00:00
22873	2022-03-21 00:00:00	2022-03-21 00:00:00
22875	2022-03-23 00:00:00	2022-03-23 00:00:00
22878	2022-03-26 00:00:00	2022-03-26 00:00:00

3. **On which date are the shots fired correlating with the incident report date?**

Select shots_fired.incident_number, shots_fired.incident_date, shots_fired.district_cd,
from shooting left join incidents on shots_fired.incident_date =
incidents.occurred_on_date
order by shots_fired.incident_date;



```
19  Select shots_fired.incident_number,
20  shots_fired.incident_date,
21  shots_fired.district_cd
22  from shots_fired left join incidents on shots_fired.incident_date = incidents.occurred_on_date
23  order by shots_fired.incident_date;
24
```

100% 40:22

Result Grid | Filter Rows: Search | Export:

incident_numb...	incident_date	district_...
I152000042-00	2015-01-01 00:05:00	A15
I152000093-00	2015-01-01 04:17:00	C11
I152000514-00	2015-01-02 15:51:00	B3
I152000600-00	2015-01-02 22:25:00	E13
I152000856-00	2015-01-03 23:57:00	B2
I152000872-00	2015-01-04 00:13:00	B2
I152000965-00	2015-01-04 15:30:00	B2
I152001379-00	2015-01-05 22:00:00	C6
I152001625-00	2015-01-06 19:39:00	B3
I152001947-00	2015-01-08 12:30:00	B3
I152002240-00	2015-01-09 02:09:00	A1

4. **What are the top 10 districts that have the most incident reports?**

```
Select count(incidents.incident_number) as total_incidents,
incidents.district,
districts.district_name
from incidents
inner join districts
on incidents.district = districts.district_cd
group by incidents.district
order by total_incidents desc
limit 10;
```

```
25  Select count(incidents.incident_number) as total_incidents,
26  incidents.district,
27  districts.district_name
28  from incidents
29  inner join districts
30
```

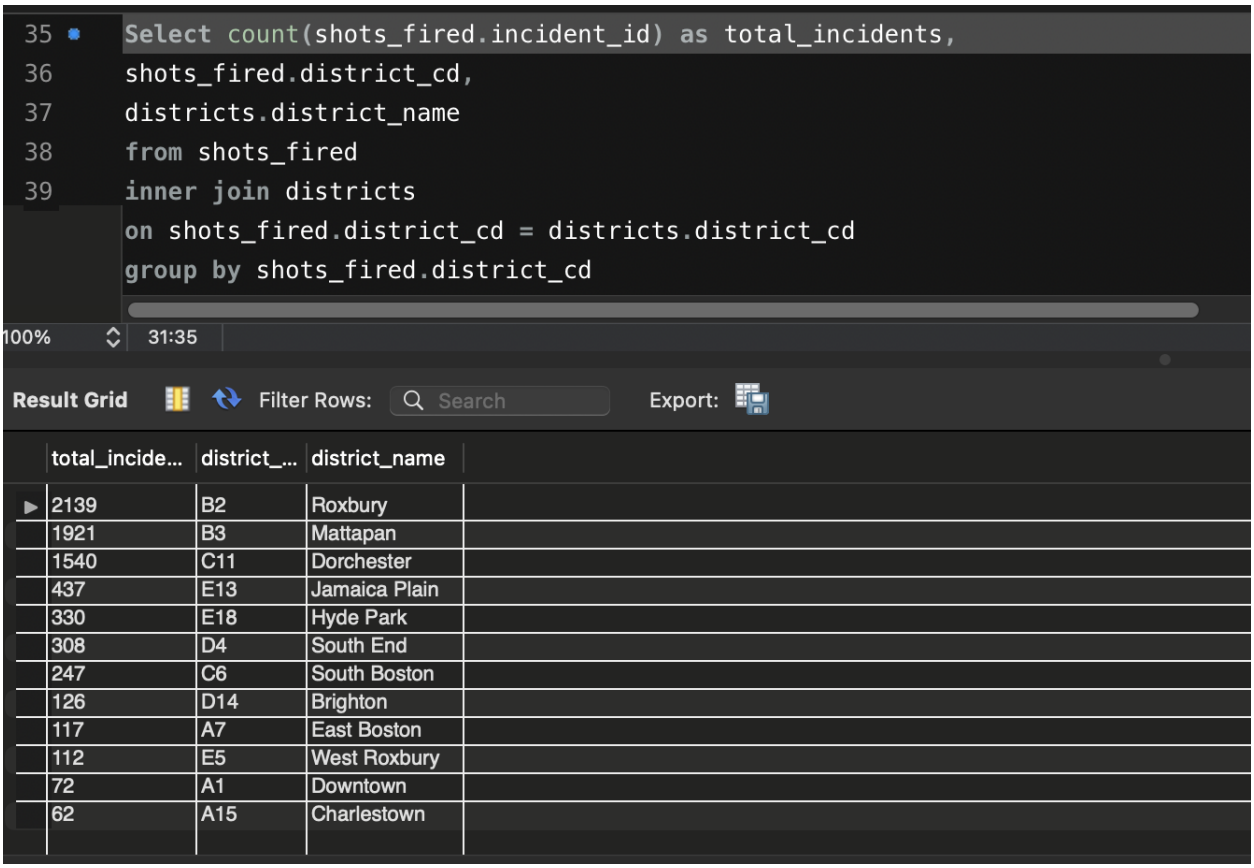
100% 18:26

Result Grid | Filter Rows: Search | Export: | Fetch rows:

total_incide...	district	district_name
30	B3	Mattapan
24	C11	Dorchester
23	B2	Roxbury
11	D4	South End
7	E18	Hyde Park
6	C6	South Boston
4	A7	East Boston
4	E13	Jamaica Plain
2	A15	Charlestown
1	A1	Downtown

5. **In which districts were each shot fired?**

Select count(shots_fired.incident_id) as total_incidents,
shots_fired.district_cd,
districts.district_name,
from shots_fired
inner join districts
on shots_fired.district_cd = districts.district_cd
group by shots_fired.district_cd
order by total_incidents desc;



```
35  Select count(shots_fired.incident_id) as total_incidents,
36  shots_fired.district_cd,
37  districts.district_name
38  from shots_fired
39  inner join districts
    on shots_fired.district_cd = districts.district_cd
    group by shots_fired.district_cd
```

100% 31:35

Result Grid Filter Rows: Search Export:

total_incide...	district_...	district_name
2139	B2	Roxbury
1921	B3	Mattapan
1540	C11	Dorchester
437	E13	Jamaica Plain
330	E18	Hyde Park
308	D4	South End
247	C6	South Boston
126	D14	Brighton
117	A7	East Boston
112	E5	West Roxbury
72	A1	Downtown
62	A15	Charlestown

6. **In which districts have the most shootings occurred?**

Select count(shooting.incident_number) as total_incidents,
shooting.district_cd,
districts.district_name,
districts.phone
from shooting
join as shooting.district_cd = districts.district_cd
group by district_cd
order by total_incidents;

```
44  Select count(shooting.incident_number) as total_incidents,
45  shooting.district_cd,
46  districts.district_name,
47  districts.phone
48  from shooting
49  join districts on shooting.district_cd = districts.district_cd
50  group by district_cd
51
```

100% 20:46

Result Grid Filter Rows: Search Export:

total_incide...	district_...	district_name	phone
1	D14	Brighton	(617) 343-4260
1	E5	West Roxbury	(617) 343-4560
1	A1	Downtown	(617) 343-4240
2	A15	Charlestown	(617) 343-4888
4	A7	East Boston	(617) 343-4220
5	E13	Jamaica Plain	(617) 343-5630
7	C6	South Boston	(617) 343-4730
7	E18	Hyde Park	(617) 343-5600
8	D4	South End	(617) 343-4250
22	C11	Dorchester	(617) 343-4330
24	B2	Roxbury	(617) 343-4270
33	B3	Mattapan	(617) 343-4700

7. **On what date was the number of guns recovered the highest?**

```
Select collection_date,
(crimeguns_recovered +guns_recovered +buybackguns_recovered) as
total_guns_recovered
from firearm_recovery
order by total_guns_recovered
```

```
53  Select collection_date,
54  (crimeguns_recovered + guns_recovered + buybackguns_recovered) as total_guns_recovered
55  from firearm_recovery
56  order by total_guns_recovered;
57
58  SELECT shooting.incident_number,
```

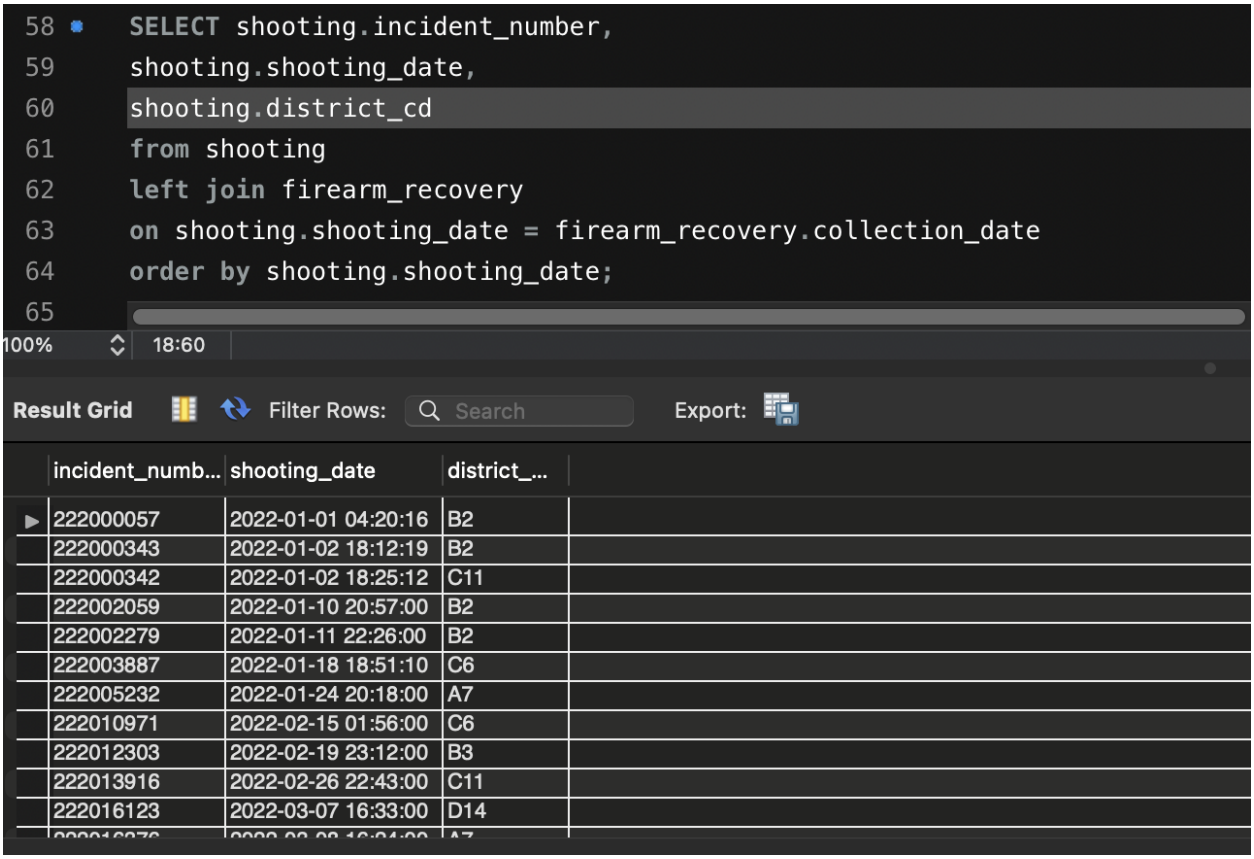
100% 31:56

Result Grid Filter Rows: Search Export:

collection_date	total_guns_recover...
2014-11-03 00:00:00	NULL
2015-05-21 00:00:00	NULL
2017-03-28 00:00:00	NULL
2017-05-01 00:00:00	NULL
2017-08-02 00:00:00	NULL
2017-08-30 00:00:00	NULL
2017-11-14 00:00:00	NULL
2018-01-22 00:00:00	NULL
2018-01-24 00:00:00	NULL
2018-04-23 00:00:00	NULL
2018-04-27 00:00:00	NULL
2018-05-16 00:00:00	NULL
2018-05-21 00:00:00	NULL

8. **Were firearms recovered on the dates when shootings occurred?**

```
SELECT shooting.incident_number,
shooting.shooting_date,
shooting.district_cd
from shooting
left join firearm_recovery
on shooting.shooting_date = firearm_recovery.collection_date
order by shooting.shooting_date;
```



```
58  SELECT shooting.incident_number,
59  shooting.shooting_date,
60  shooting.district_cd
61  from shooting
62  left join firearm_recovery
63  on shooting.shooting_date = firearm_recovery.collection_date
64  order by shooting.shooting_date;
65
```

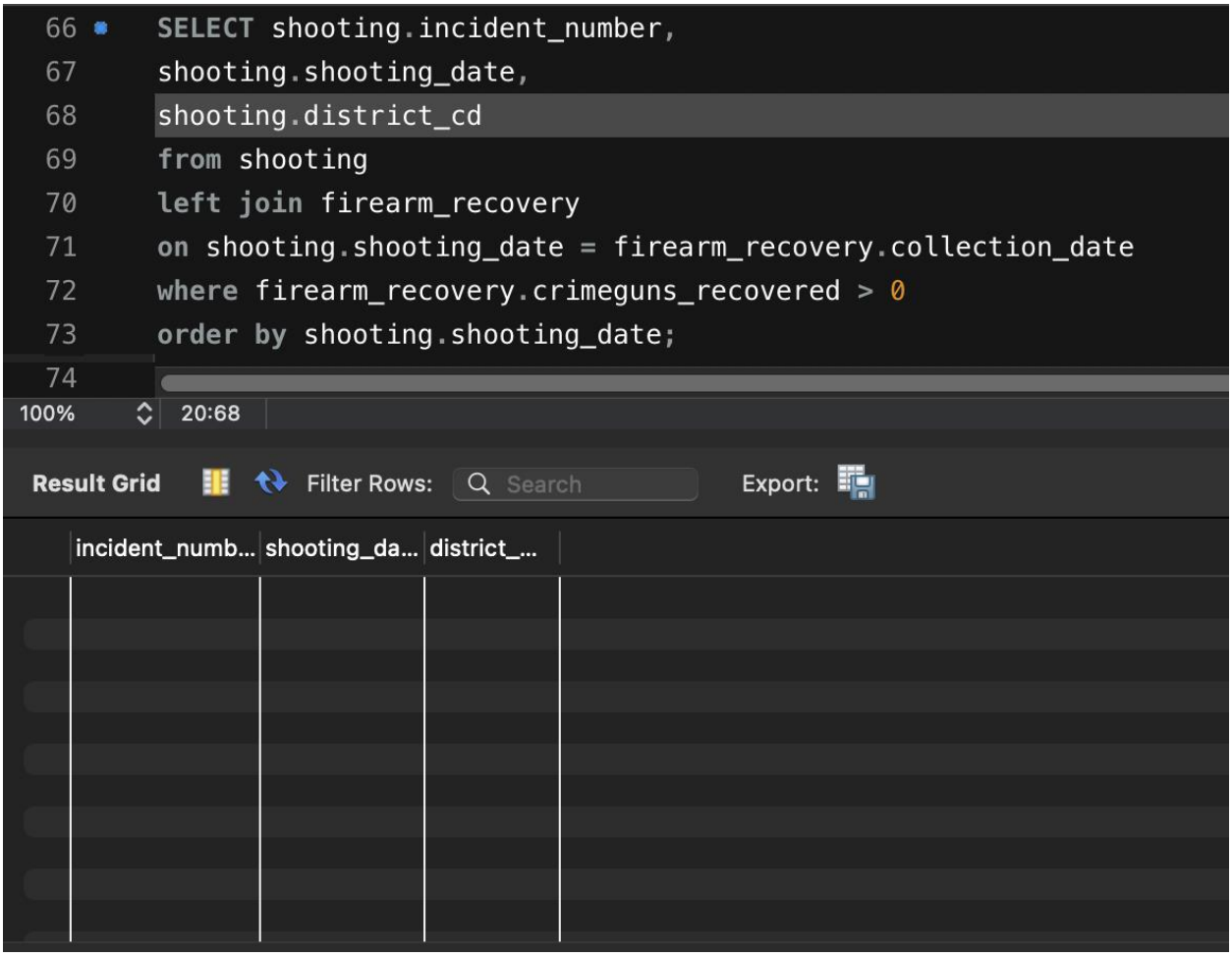
100% 18:60

Result Grid | Filter Rows: Search | Export:

incident_numb...	shooting_date	district_...
222000057	2022-01-01 04:20:16	B2
222000343	2022-01-02 18:12:19	B2
222000342	2022-01-02 18:25:12	C11
222002059	2022-01-10 20:57:00	B2
222002279	2022-01-11 22:26:00	B2
222003887	2022-01-18 18:51:10	C6
222005232	2022-01-24 20:18:00	A7
222010971	2022-02-15 01:56:00	C6
222012303	2022-02-19 23:12:00	B3
222013916	2022-02-26 22:43:00	C11
222016123	2022-03-07 16:33:00	D14

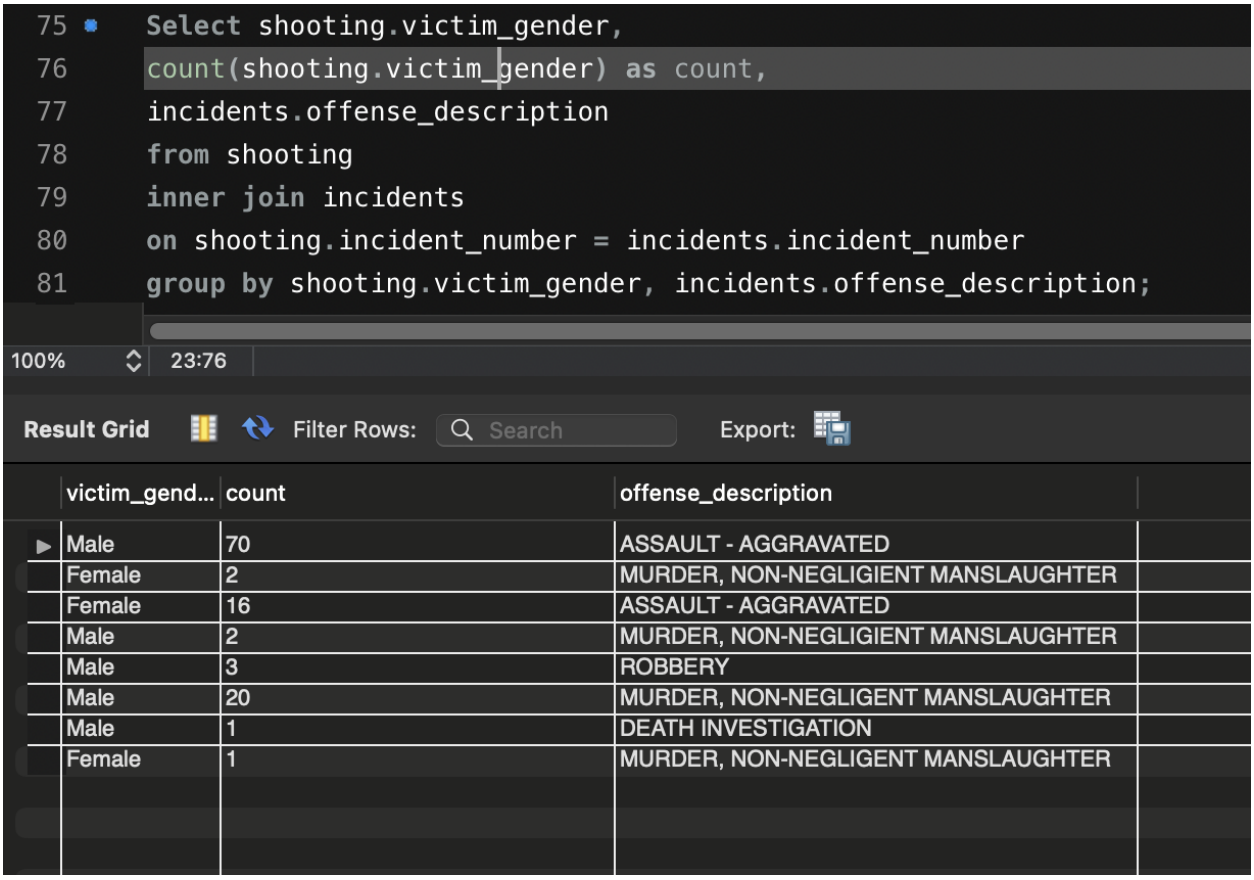
9. **How many crime-related firearms were collected on the dates when shootings occurred?**

```
SELECT shooting.incident_number,
shooting.shooting_date,
shooting.district_cd
from shooting
left join firearm_recovery
on shooting.shooting_date = firearm_recovery.collection_date
where firearm_recovery.crimeguns_recovered > 0
order by shooting.shooting_date;
```



10. Which gender was involved with the most shooting?

```
Select shooting.victim_gender,
count(shooting.victim_gender) as count,
incidents.offense_description
from shooting
inner join incidents
on shooting.incident_number = incidents.incident_number
group by shooting.victim_gender, incidents.offense_description;
```



```
75    Select shooting.victim_gender,
76    count(shooting.victim_gender) as count,
77    incidents.offense_description
78    from shooting
79    inner join incidents
80    on shooting.incident_number = incidents.incident_number
81    group by shooting.victim_gender, incidents.offense_description;
```

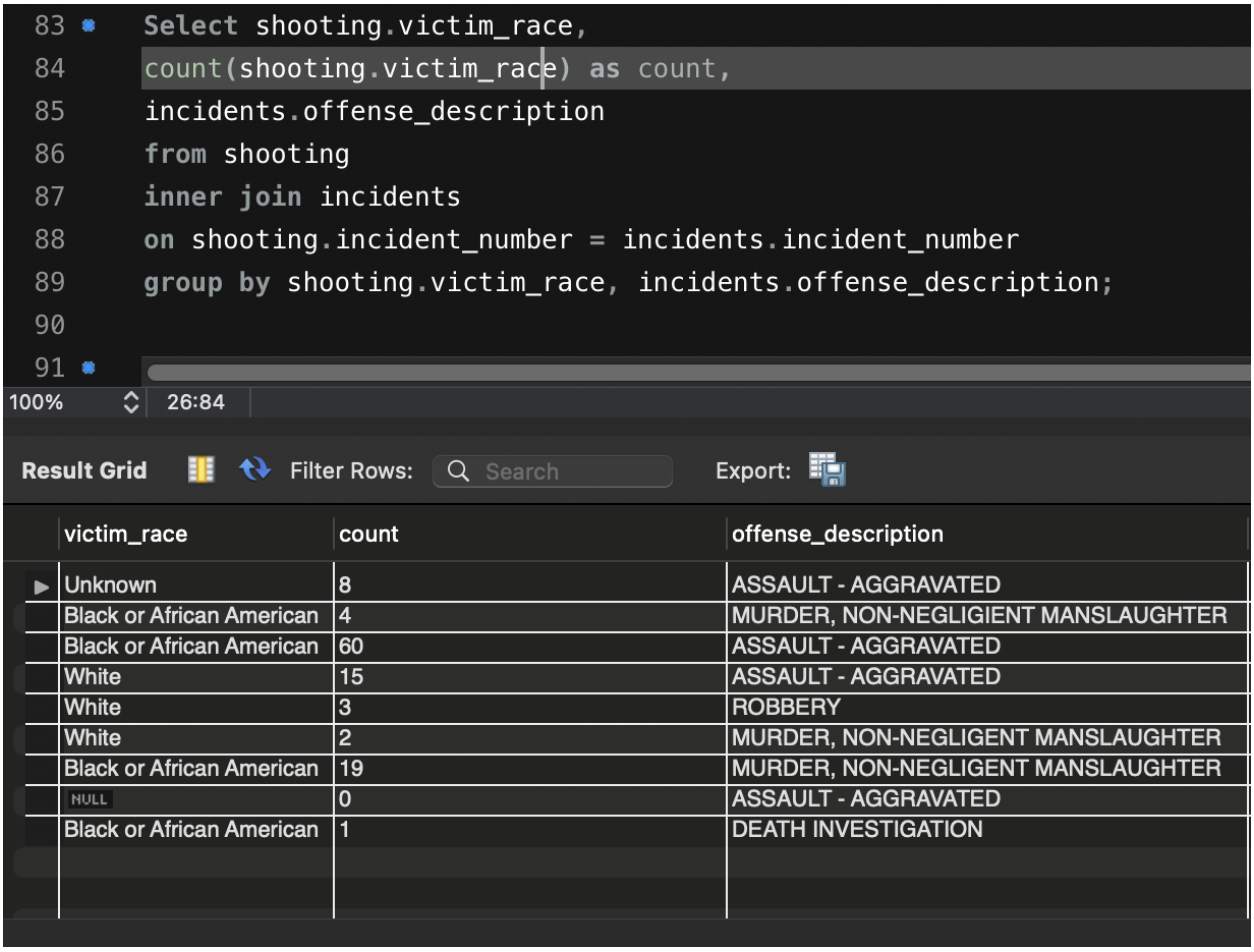
100% 23:76

Result Grid Filter Rows: Search Export:

victim_gend...	count	offense_description
Male	70	ASSAULT - AGGRAVATED
Female	2	MURDER, NON-NEGLIGIENT MANSLAUGHTER
Female	16	ASSAULT - AGGRAVATED
Male	2	MURDER, NON-NEGLIGIENT MANSLAUGHTER
Male	3	ROBBERY
Male	20	MURDER, NON-NEGLIGENT MANSLAUGHTER
Male	1	DEATH INVESTIGATION
Female	1	MURDER, NON-NEGLIGENT MANSLAUGHTER

11. Which race was involved with the most shooting?

```
Select shooting.victim_race,
count(shooting.victim_race) as count,
incidents.offense_description
from shooting
inner join incidents
on shooting.incident_number = incidents.incident_number
group by shooting.victim_race, incidents.offense_description;
```



victim_race	count	offense_description
Unknown	8	ASSAULT - AGGRAVATED
Black or African American	4	MURDER, NON-NEGLIGIENT MANSLAUGHTER
Black or African American	60	ASSAULT - AGGRAVATED
White	15	ASSAULT - AGGRAVATED
White	3	ROBBERY
White	2	MURDER, NON-NEGLIGENT MANSLAUGHTER
Black or African American	19	MURDER, NON-NEGLIGENT MANSLAUGHTER
NULL	0	ASSAULT - AGGRAVATED
Black or African American	1	DEATH INVESTIGATION

12. Which ethnicity was involved with the most shooting?

```
Select shooting.victim_ethnicity,
count(shooting.victim_ethnicity) as count,
incidents.offense_description
from shooting
inner join incidents
on shooting.incident_number = incidents.incident_number
group by shooting.victim_ethnicity, incidents.offense_description;
```

```
91  Select shooting.victim_ethnicity,
92  count(shooting.victim_ethnicity) as count,
93  incidents.offense_description
94  from shooting
95  inner join incidents
96  on shooting.incident_number = incidents.incident_number
97  group by shooting.victim_ethnicity, incidents.offense_description;
98
```

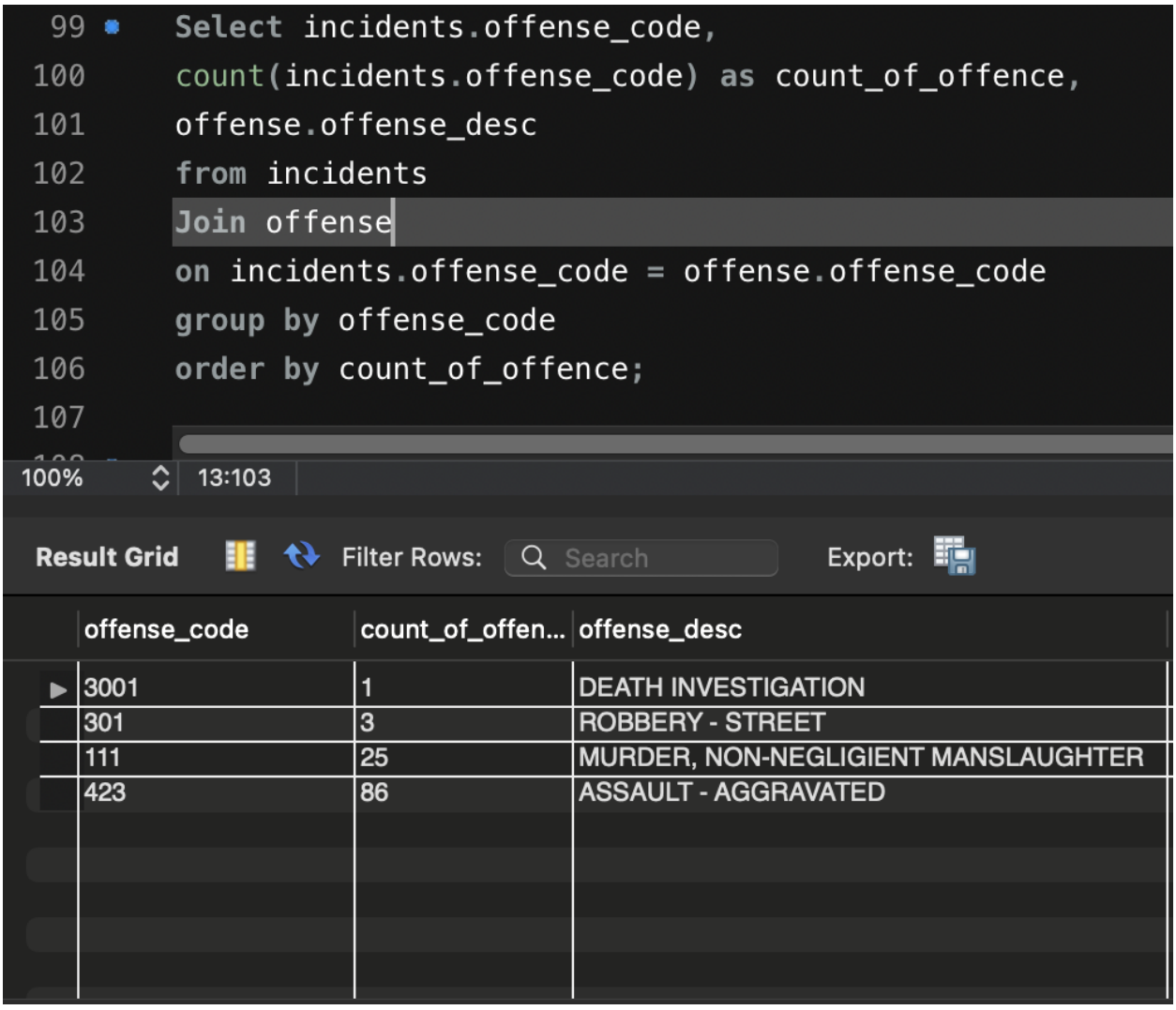
100% 26:92

Result Grid Filter Rows: Search Export:

victim_ethnicity	count	offense_description
NULL	0	MURDER, NON-NEGLIGIENT MANSLAUGHTER
Not Hispanic or Latinx	45	ASSAULT - AGGRAVATED
Unknown	11	ASSAULT - AGGRAVATED
NULL	0	ASSAULT - AGGRAVATED
Not Hispanic or Latinx	3	MURDER, NON-NEGLIGIENT MANSLAUGHTER
Not Hispanic or Latinx	1	ROBBERY
Hispanic or Latinx	3	MURDER, NON-NEGLIGENT MANSLAUGHTER
Unknown	3	MURDER, NON-NEGLIGENT MANSLAUGHTER
Not Hispanic or Latinx	15	MURDER, NON-NEGLIGENT MANSLAUGHTER
Hispanic or Latinx	2	ROBBERY
Not Hispanic or Latinx	1	DEATH INVESTIGATION

13. Which type of offense is the most recorded?

```
Select incidents.offense_code,
count(incidents.offense_code) as count_of_offence,
offense.offense_desc
from incidents
Join offense
on incidents.offense_code = offense.offense_code
group by offense_code
order by count_of_offence;
```



```
99  Select incidents.offense_code,
100 count(incidents.offense_code) as count_of_offence,
101 offense.offense_desc
102 from incidents
103 Join offense
104 on incidents.offense_code = offense.offense_code
105 group by offense_code
106 order by count_of_offence;
107
```

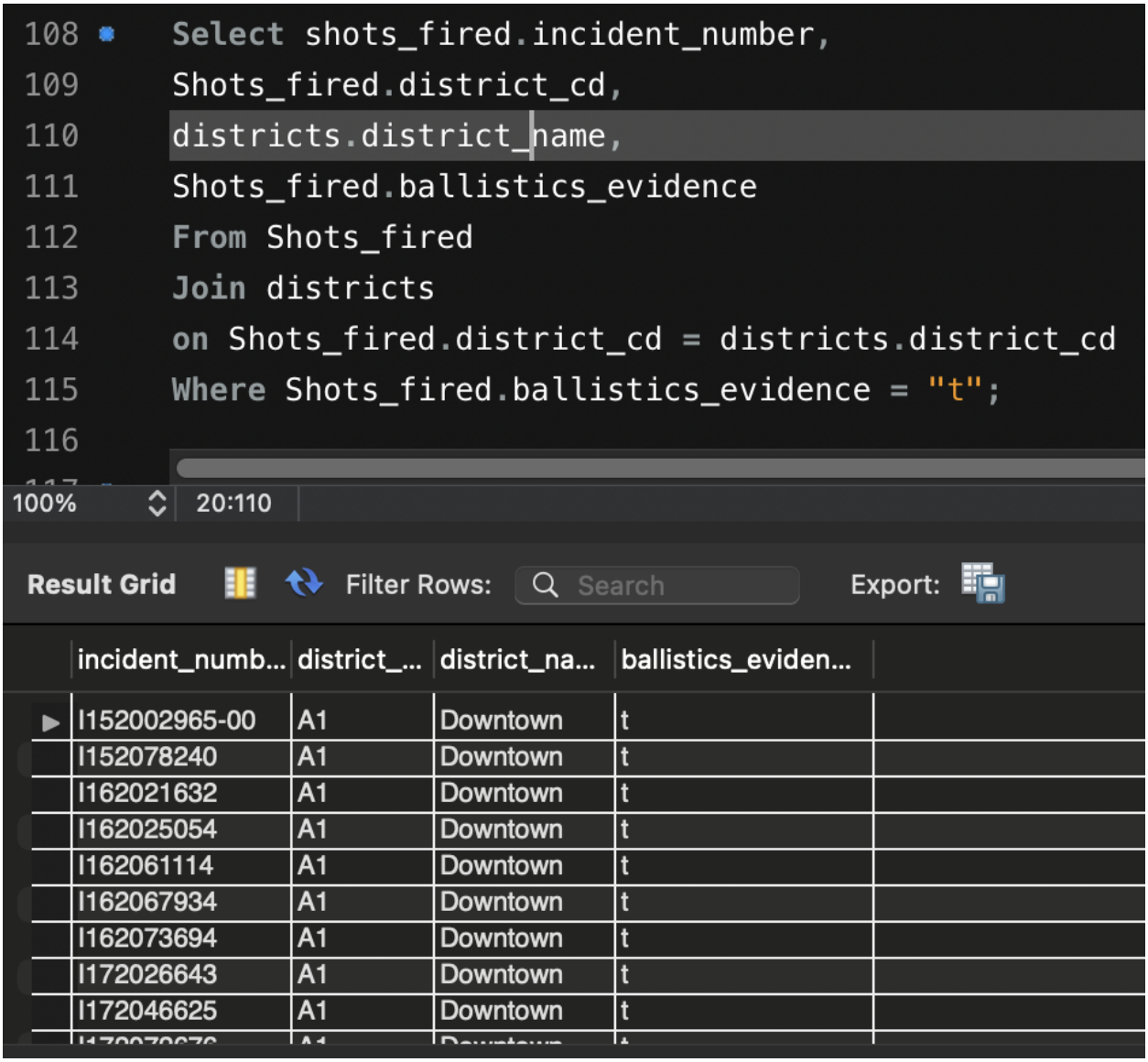
100% 13:103

Result Grid Filter Rows: Search Export:

offense_code	count_of_offen...	offense_desc
3001	1	DEATH INVESTIGATION
301	3	ROBBERY - STREET
111	25	MURDER, NON-NEGLIGIENT MANSLAUGHTER
423	86	ASSAULT - AGGRAVATED

14. How many incidents had ballistic evidence recovered?

```
Select shots_fired.incident_number,
Shots_fired.district_cd,
districts.district_name,
Shots_fired.ballistics_evidence
From Shots_fired
Join districts
on Shots_fired.district_cd = districts.district_cd
Where Shots_fired.ballistics_evidence = "t";
```



```
108  Select shots_fired.incident_number,
109  Shots_fired.district_cd,
110  districts.district_name,
111  Shots_fired.ballistics_evidence
112  From Shots_fired
113  Join districts
114  on Shots_fired.district_cd = districts.district_cd
115  Where Shots_fired.ballistics_evidence = "t";
116
```

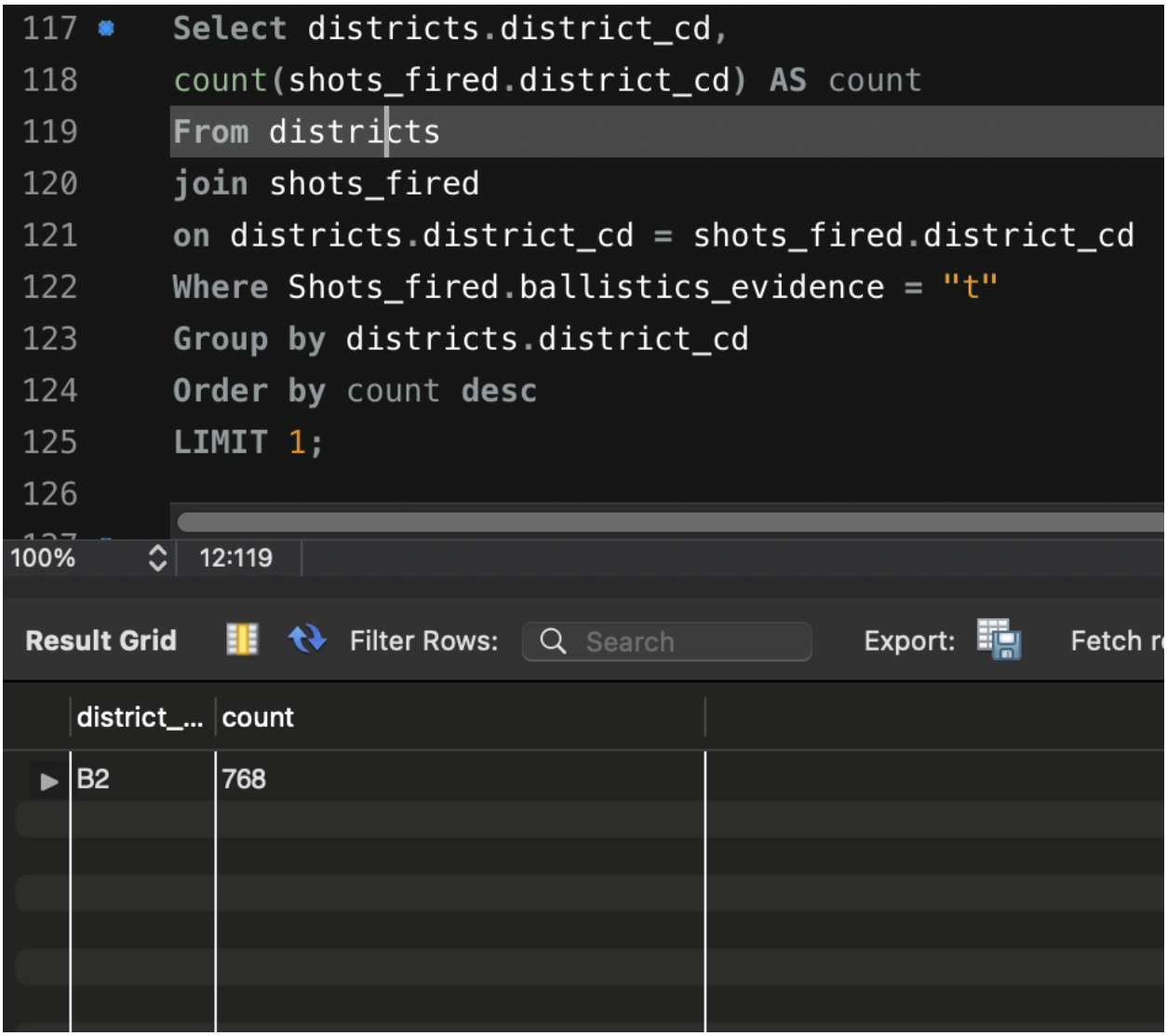
100% 20:110

Result Grid Filter Rows: Search Export:

incident_numb...	district_...	district_na...	ballistics_eviden...
I152002965-00	A1	Downtown	t
I152078240	A1	Downtown	t
I162021632	A1	Downtown	t
I162025054	A1	Downtown	t
I162061114	A1	Downtown	t
I162067934	A1	Downtown	t
I162073694	A1	Downtown	t
I172026643	A1	Downtown	t
I172046625	A1	Downtown	t

15. Which districts had the most ballistic evidence recovered?

```
Select districts.district_cd,
count(shots_fired.district_cd) AS count
From districts
join shots_fired
on districts.district_cd = shots_fired.district_cd
Where Shots_fired.ballistics_evidence = "t"
Group by districts.district_cd
Order by count desc
LIMIT 1;
```



16. Which incidents resulted in multiple victims?

Select incidents.incident_number,
Shooting.multi_victim
From incidents inner join shooting
on incidents.incident_number = shooting.incident_number
Where shooting.multi_victim = "t"
Group by incidents.incident_number,
shooting.multi_victim;

```
127    Select incidents.incident_number,
128    Shooting.multi_victim
129    From incidents inner join shooting
130    on incidents.incident_number = shooting.incident_number
131    Where shooting.multi_victim = "t"
132    Group by incidents.incident_number,
133    shooting.multi_victim;
134
```

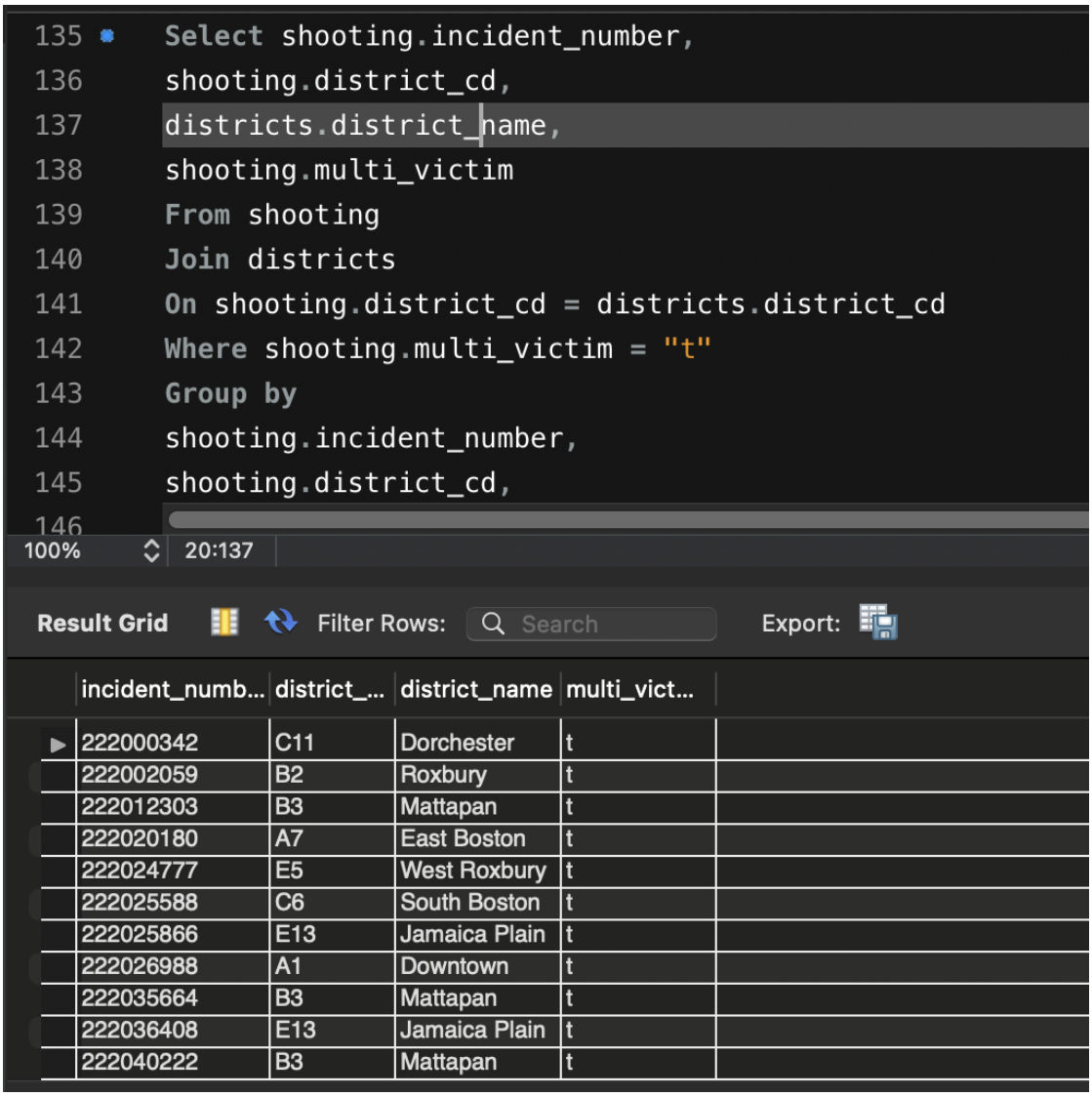
100% 14:130

Result Grid Filter Rows: Search Export:

incident_numb...	multi_vict...
222012303	t
222020180	t
222024777	t
222025588	t
222025866	t
222026988	t
222035664	t
222036408	t
222040222	t
222049436	t
222049446	t

17. Which district had more than 1 victim?

```
Select shooting.incident_number,
shooting.district_cd,
districts.district_name,
shooting.multi_victim
From shooting
Join districts
On shooting.district_cd = districts.district_cd
Where shooting.multi_victim = "t"
Group by
shooting.incident_number,
shooting.district_cd,
districts.district_name,
shooting.multi_victim;
```



```
135  Select shooting.incident_number,
136  shooting.district_cd,
137  districts.district_name,
138  shooting.multi_victim
139  From shooting
140  Join districts
141  On shooting.district_cd = districts.district_cd
142  Where shooting.multi_victim = "t"
143  Group by
144  shooting.incident_number,
145  shooting.district_cd,
146
```

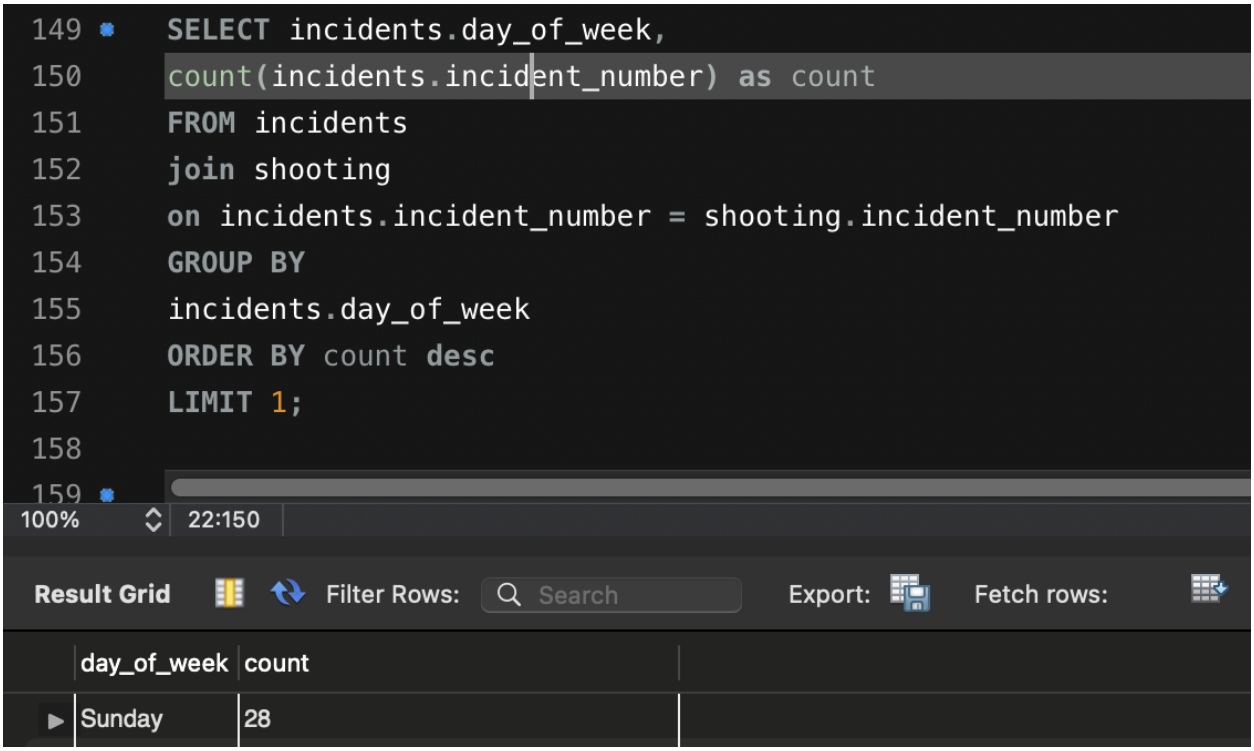
100% 20:137

Result Grid Filter Rows: Search Export:

incident_numb...	district_...	district_name	multi_vict...
222000342	C11	Dorchester	t
222002059	B2	Roxbury	t
222012303	B3	Mattapan	t
222020180	A7	East Boston	t
222024777	E5	West Roxbury	t
222025588	C6	South Boston	t
222025866	E13	Jamaica Plain	t
222026988	A1	Downtown	t
222035664	B3	Mattapan	t
222036408	E13	Jamaica Plain	t
222040222	B3	Mattapan	t

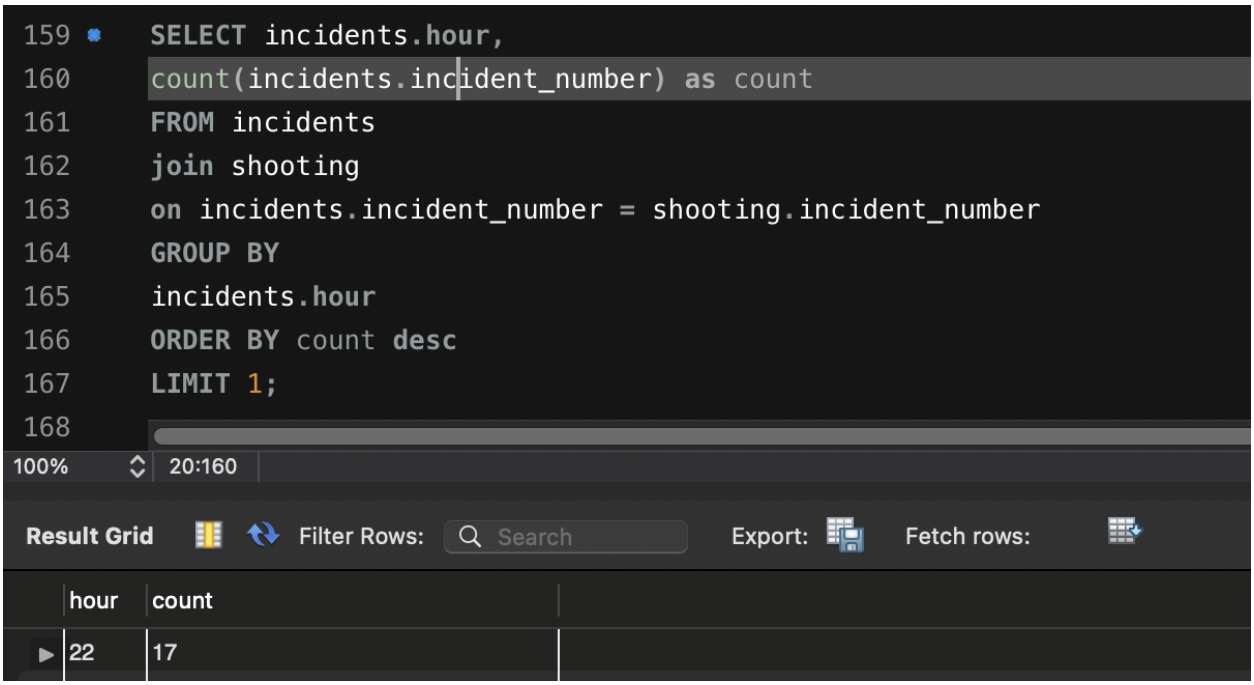
18. Which day of week had the least shooting?

```
SELECT Incidents.day_of_week,
count(incidents.incident_number) as count
FROM incidents
join shooting
on incidents.incident_number = shooting.incident_number
GROUP BY
incidents.day_of_week
ORDER BY count desc
LIMIT 1;
```



19. Which hour of the day has the most shootings?

```
SELECT incidents.hour,
count(incidents.incident_number) as count
FROM incidents
join shooting
on incidents.incident_number = shooting.incident_number
GROUP BY
incidents.hour
ORDER BY count desc
LIMIT 1;
```



20. What is the highest number of buyback guns recovered for a particular incident date?

```
SELECT firearm_recovery.collection_date,
firearm_recovery.buybackguns_recovered
FROM firearm_recovery
INNER JOIN incidents
ON firearm_recovery.collection_date = incidents.occurred_on_date
GROUP BY
firearm_recovery.collection_date,
firearm_recovery.buybackguns_recovered
ORDER BY firearm_recovery.buybackguns_recovered;
```